10.5 A METHOD TO INFER HISTORIC TORNADO FREQUENCY FROM RADIOSONDE RECORDS

Matthew J. Menne*
NOAA/NESDIS/NCDC, Asheville, North Carolina

1. INTRODUCTION

The National Oceanic and Atmospheric Administration's Storm Prediction Center has maintained a storm event database that includes observations of tornado events in the United States with a period of record that begins in 1950. Counts of the annual total number of tornadoes recorded in the database suggest an almost linear increase in tornado frequency, especially in weaker tornadoes. Annual totals increase from about 200 tornadoes in 1950 to about 1000 by the early 1990s (Fig. 1). The frequency of tornado occurrence as recorded in the database, however, is thought to be significantly impacted by inhomogeneities present in the archive.

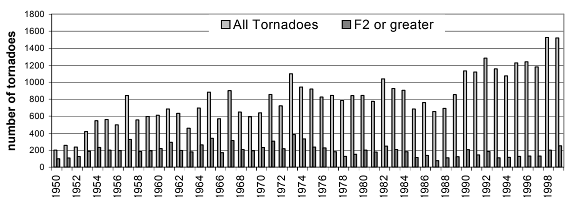

Figure 1. Annual number of tornadoes from the Storm Prediction Center's storm event data base.

Inhomogeneities in storm event records arise from changes in observing practices associated with, for example, the introduction of radar and changes in institutional requirements for forecast verification. Such changes have led to improvements in tornado detection and verification, but their impact on the historical record of storm events hampers efforts to quantify the true frequency of tornado occurrence through time. Given that establishing counts of truly unique tornadoes also can be problematic (Doswell and Burgess, 1988), tornado-day measures have sometimes been used to help avoid potential non-climatic artifacts in the observational record. A tornado day is simply a day in which at least one tornado was observed. Fig. 2 depicts a measure of the annual number of tornado days in the central United Stations. The figure is based on the number of times one or more tornadoes was reported within 200 km of a radiosonde station around the time of the 0000 UTC balloon release. As is evident from the figure, trends can be present even in some tornado-day measures, which are probably affected to some degree by the same inhomogeneities as raw tornado counts.

* Corresponding author address: Matthew J. Menne, National Climatic Data Center, Climate Analysis Branch, Asheville, North Carolina, 28801; e-mail: Matthew.Menne@noaa.gov

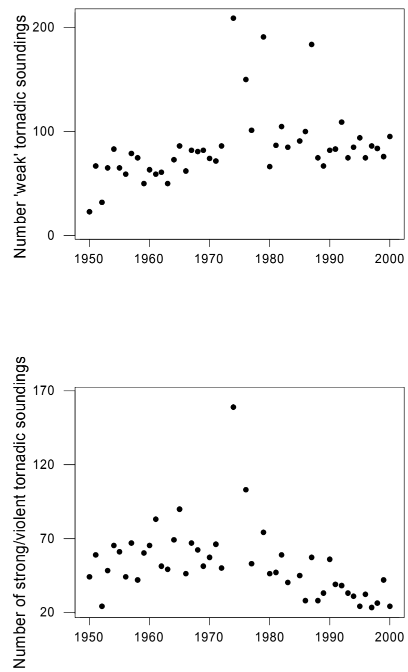

Figure 2. Observed number of weak and significant tornado soundings in the central United States that meet the 0000 UTC proximity criteria described in section 3.

In previous studies (e.g., Brooks et. al, 1994; Rasmussen and Blanchard, 1998), investigators have compared the frequency distributions of samples of soundings associated with tornadic and non-tornadic thunderstorms with the aim of identifying differences in severe local storm environments for use in forecasting and for comparison with model simulations. Recent work by Rasmussen and Blanchard (1998) and Craven (2001), linking observed storm events and various sounding-based parameters, revealed some apparent differences between environments classified as tornadic and non-tornadic in, for example, values of low level shear and height of the Lifting Condensation Level (LCL). Rasmussen and Blanchard (1998) examined all 0000 UTC soundings for the United States for the year 1992 while Craven (2001) analyzed soundings for the years 1997 through 2000.

In this study, a similar comparison of storm environments is presented using large samples of soundings and storm events. Multiple logistic regression, in which the dependent variable is a probability, is used to estimate the likelihood of tornado occurrence given the presence of a specific set of atmospheric conditions. Observations from the most recent years in the period of record, a period when severe local storm event observations are considered to be reasonably stable through time, are used to develop the regression relationships and to quantify storm inference skill. An example is provided where the regression relationships are used to infer tornado-day occurrence for earlier years in the observational record in order to produce a more homogeneous record. This paper reports on some preliminary conclusions of the investigation.

2. DATA

Observations of severe local storm events and twice daily radiosonde observations were used in the analysis. Storm event observations were obtained from the Storm Prediction Center's storm event database and from the National Climatic Data Center's *Storm Data* database. These databases include observations of tornadoes beginning in 1950 and thunderstorm winds and hail beginning in 1955. Information on the time and location of occurrence for each storm event is included in the database.

Radiosonde observations for the period 1960 to 2000 were extracted from the *Radiosonde Data of North America* dataset produced by NOAA's Forecast System Lab and the National Climatic Data Center, available on compact disc. An array of sounding-based parameters was calculated using each twice daily (0000 and 1200 UTC) observation from a sample of 25 stations located in the central United States. Many of the commonly used sounding-based severe weather forecast parameters were calculated including Convective Available Potential Energy (CAPE), convective inhibition and measures of wind shear magnitude for various atmospheric layers.

During the 1960 - 2000 period, many radiosonde stations were subject to changes in location. In most cases, a station move was not accompanied by the assignment of a new station identification number. However, in some cases, a new identifier was assigned and the former station was considered decommissioned. All station decommissions in the sample of stations used in this study, however, were linked with the coincident commissioning of a replacement station relatively nearby. In these cases, in order to form a continuous time series of radiosonde observations for the study period, observations from the two stations were merged and treated as one station. The linked stations are mapped with an arbitrary identification number in Fig. 3. Station moves not involving the assignment of a new station identifier also are shown. Many of the station moves involving a station decommissioning occurred in the mid-1990s.

3. PROCEDURE

Each 0000 and 1200 UTC radiosonde observation was associated with one of six different categories of storm event type using the "proximity method." The proximity method (Darkow, 1969; Brooks et al., 1994) is simply one in which a storm event in the vicinity of a radiosonde station is linked to a radiosonde observation when the event occurred within specified space and time proximity criteria. The assumption behind this method is that the sounding is representative of the environment in which the storm formed. In this study, a storm event had to occur with 200 km of a radiosonde station and within the period two hours before to four hours after the 0000 or 1200 UTC balloon release for the sounding to be assigned to that storm event class.

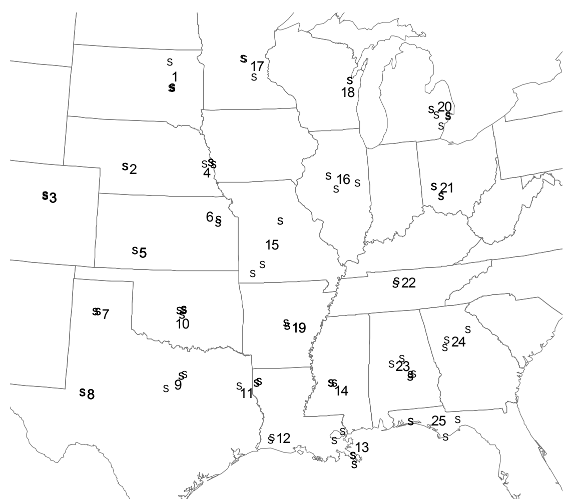

Figure 3. Locations of radiosonde stations used in the study.

Storm events were classified as one of six event types according to the criteria outlined in Table 1. This classification is similar to those used by Mead (1997), Rasmussen and Blanchard (1998) and Craven (2001). In these classifications, an attempt is made to distinguish between two types of supercell thunderstorms: tornadic (F2 or greater tornado) and non-tornadic (large hail but no significant tornado). In earlier studies only F2 or greater Fujita scale tornadoes were used to mark the tornadic supercell class since the parent thunderstorm of stronger tornadoes is likely to be a supercell. Using the F2 or greater class also avoids the potential misattribution of damage as having been caused by weak tornadoes rather than straight-line winds (Doswell and Burgess, 1988). While there is no guarantee that supercells likewise produce all large hail observations, Mead (1997) and others justify using such reports to classify non-tornadic supercells based on observational experience that suggests that there is a high probability of mid-level updraft rotation coincident with such reports.

Table 1. Classification of storm event type

Storm Type	Definition
1	small hail (<0.75" diameter) or weak thunderstorm wind (<50 knts)
2	moderate hail (0.75 to 1.99" diameter) or thunderstorm wind (50 to 65 knts)
3	weak tornado, F0 or F1
4	severe thunderstorm wind (>65 knts)
5	hail > 2.00" diameter (non-tornadic supercell) and no reported tornado
6	strong or violent tornado, F2 or greater (tornadic supercell)
-999	no storm event reported within the proximity criteria

4. RESULTS

Fig. 4 shows a box plot of the surface to 1 km Above Ground Level (AGL) wind vector difference and Fig. 5 depicts the mean level of the LCL height AGL for a sample of proximity soundings that includes all 0000 UTC observations for the period 1991-2000 from the network of observing stations shown in Fig. 3. Results based on this sample, stratified by the classification shown in Table 1, are similar to those shown by Rasmussen and Blanchard (1998) and Craven (2001).

While these and other sounding-based parameters, when viewed en masse using observations from a number of stations, appear to show some evidence of differentiation by storm type, it should be noted that there is station dependence present in many parameter frequency distributions. This is especially true for those parameters sensitive to boundary layer moisture content (e.g., CAPE and LCL height). Consequently, a moderately high LCL height observation for a given month at a station east of the Mississippi River may be considered quite low at a station west of the Mississippi River. In some extreme cases, the mean value for the group may never be attained at a member station. This is true even for proximity soundings assigned to the same storm event class. An example of such station dependency is provided in Fig. 6. The figure suggests that LCL height AGL values for stations in the northern and southern plains are generally higher than those east of the Mississippi River and along the Gulf Coast even under significant tornadic conditions. The presence of station dependency complicates attempts to identify differences in severe weather environments.

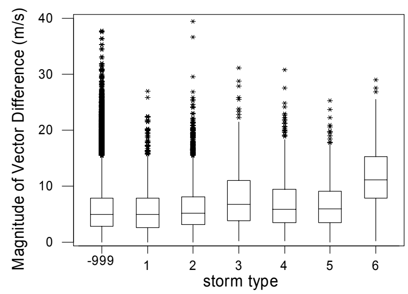

Figure 4. Box plots of the magnitude of the vector difference between the surface and 1 km AGL wind.

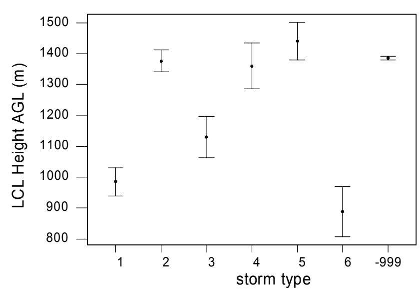

Figure 5. Mean and 95% confidence limits for all classified 0000 UTC soundings for the years 1991-2000 from the 25 locations shown in Fig. 3.

A practical solution to the station dependency issue is to standardize parameter values at each station. Unfortunately, the frequency distributions of many of the sounding-based parameters are not normal since many of the parameters are zero-bounded. In fact, the most frequent magnitude of CAPE at many locations in some months is zero. In winter months at the more northerly stations, the probability of observing zero CAPE approaches 1.0. Consequently, a seasonal dependence may also be present in some parameter distributions.

Given that expressing observations as *z*-scores is not reasonable for many parameters and considering the potential for a seasonal dependency, expressing values as station-based monthly probability quantiles was explored as a means to standardize the frequency distribution. Fig. 7 shows box plots of a three-parameter gamma-fit probability quantile distribution for LCL height for all type 6 (F2 or greater tornado) proximity soundings at each station. In many cases, obvious station differences in the nature of frequency distributions remain. That this is true can be illustrated again using the probability of observing a value of zero for CAPE. In May, for example, moderate values of CAPE (1000 – 3000 J/m-2) are associated with frequencies around the 75th percentile in the southern plains. In the northern plains, the probability of observing zero CAPE in May approaches 70%. In spite of these regional differences, nearby stations may be characterized by similar frequency distributions.

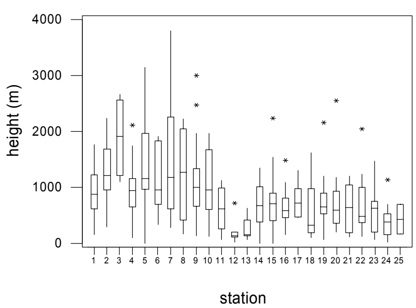

Figure 6. Box plots showing the frequency distribution, by station, of LCL height AGL for all 0000 UTC sounding classified as storm type 6 (tornadic supercell).

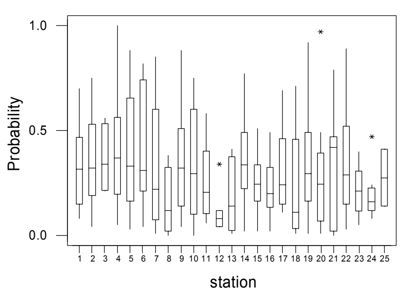

Figure 7. As in Fig. 5, except in this case box plots show the distribution of frequency quantiles for LCL height AGL. Frequency quantiles were estimated by fitting a three-parameter gamma distribution for each month and for each station.

Since analysis at the station level is severely limited by the small number of observations for some storm classes, a regional approach may be required. Consequently, the results of a simple regionalization are discussed below.

A cluster analysis was performed based on five vectors of monthly parameter-based statistics from each of the 25 stations. Monthly mean statistics for parameters whose frequency distribution appears to show some differentiation among the 6 storm types were used. These included monthly averages of CAPE, LCL height, surface to 1 km AGL wind shear, and mean wind speed in the surface to 6 km AGL layer. The monthly probability of zero CAPE at each station was used as the fifth vector. Clustering was based on Euclidean distance and the Ward linkage method (Wilks, 1995). The dendrogram for a six cluster solution is shown in Fig. 8.

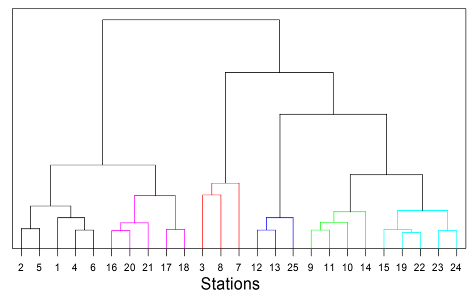

Figure 8. Dendrogram for a six cluster solution using monthly statistics for four sounding-based parameters over a base period of 1991-2000.

Using the cluster that includes stations 9, 10, 11 and 14 as an example, visual inspection of the sounding-based parameter frequency distributions suggests that the simple regionalization described by the dendrogram shown in Fig. 8 has grouped stations with similar distributions. Hosking and Wallis (1997) present a methodology to formally assess the regional selections based on the L-moments of distribution of observations from the region's member stations. Results from such an assessment will be discussed in a forthcoming report. Nevertheless, there is evidence from this clustering of stations that combining stations with similar climatologies can improve the discrimination between difference classes of severe storm environments (Fig. 9).

Using observations from stations 9, 10, 11 and 14, a logistic regression model was developed to estimate the probability of occurrence of storm type 6. The model was developed using observations from the period 1986-1994 and evaluated using observations for the period 1981-85 and1996-2000. This period was selected for model development and evaluation since there is greater confidence in verification and magnitude assessment of storm events for years after 1979. The verification period also was used to select a probability for the inference of significant tornadic soundings and for estimating the bias of the prediction.

Other investigators (e.g., Brooks et al., 1994; Rasmussen and Blanchard, 1998; Cravens 2001) have evaluated the potential use of sounding-based predictors in discriminating severe weather type trough the use of 2 x 2 contingency tables associated with the calculation of various skill scores. Skill scores combine measures such as the probability of detection and false alarm ratio to quantify various aspects of forecast success (Wilks, 1995). An attempt can be made to select a forecast parameter threshold by finding a value that optimizes a chosen skill score. This is equivalent to logistic regression in which the independent variable is also dichotomous (i.e., above or below the threshold). In multivariable logistic regression, however, each independent variable can be treated as continuous and is presented here as an alternative to the use of contingency tables. Skill score analysis then can be done to select an appropriate probability level above which an event is estimated to occur (e.g., 0.5, 0.6, 0.9) and to assess the bias of the prediction.

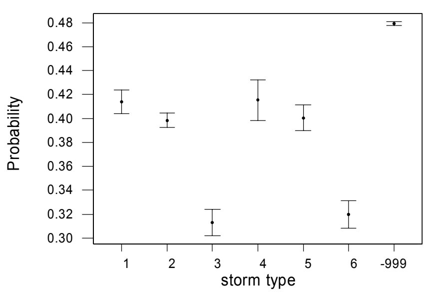

Figure 9. Mean and 95% confidence limits for the probability level of LCL height AGL associated with each of the storm types defined in Table 1. Probability distributions are based on observations from stations 9, 10, 11 and 14 for the period 1991-2000.

The initial selection of independent variables in the analysis presented below was based on fitting a univariable regression model using each sounding-based parameter as an independent variable. The significance of each independent variable coefficient was evaluated using the Wald statistic (Hosmer and Lemeshow, 2000). Each sounding-based parameter whose coefficient was significant in a univariable model was evaluated for significance in a multivariable model using the procedure described by Hosmer and Leveshow (2000). Most of the sounding-based parameters were significant in the univariate model.

The final set of predictors was selected from the group of potential predictors retained from the univariate screening as well as a number of interaction variables using a stepwise regression procedure. Interaction variables are created as the product of two primary variables. The resulting multiple logistic regression model for stations 9, 10, 11 and 14 is given as Eq. 1:

$$P = \frac{e^{9.6236-0.00631X1-0.196X2+1.61X3-0.00003X4}}{1+e^{9.6236-0.00631X1-0.196X2+1.61X3-0.00003X4}} \qquad (1)$$

where,

P = the probability of occurrence of storm type 6 which varies from 0 to 1.
X1 = the SWEAT[1] index
X2 = surface to 6 km mean wind speed
X3 = the LCL height AGL monthly probability level, and
X4 = CAPE * surface to 6 km AGL wind shear vector magnitude

The observed frequency of storm type 6 soundings for the years 1960-2000 at stations 9, 10, 11 and 14 is shown in Fig. 10. Given that there is less confidence in the observed frequency of significant tornadoes in years prior to 1979 than in later years, Eq. (1) can be used to infer the frequency of tornadoes in the early part of the record. The time series of the inferred number of significant tornadic soundings also is shown in Fig. 10. Based on these results using only radiosonde-based predictors, there is evidence that the observed frequency of strong and violent tornado days is over estimated in the period from 1960 through about the mid-1970s. Note that this is a period when the magnitude of tornado strength, based on the Fujita scale, was estimated retrospectively using indirect evidence such as newspaper accounts (Doswell and Burgess, 1988).

[1] The Severe Weather Threat or SWEAT index, used to evaluate the potential for severe weather by combining buoyancy and wind parameters into one index, was developed by the U.S. Air Force (Miller, 1972). The parameters used in the calculation include low-level moisture (850 hPa dew point temperature—TD850), instability (total-totals index—TT—sum of 850 hPa temperature and dew point minus twice the 500 hPa temperature), low and mid-level wind speeds (850 and 500 mb winds—850WSP/5000WSP), and a temperature advection factor (veering between the 850 and 500 hPa levels—WDIR850/WDIR500). The formula is as follows

SWEAT = 12[TD850] + 20(TT index - 49) + 2(WSP850) + WSP500 + 125(sin(WDIR500 - WDIR850)

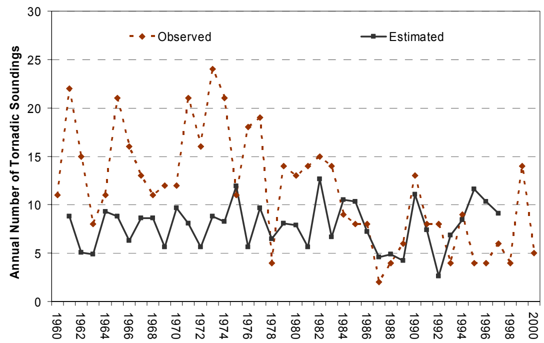

Figure 10. Number of observed and inferred type 6 soundings for stations 9, 10, 11 & 14. The number of inferred type 6 soundings was generated using Eq. (1).

5. CONCLUSIONS AND FUTURE RESEARCH

The purpose of this report is twofold: to report on an evaluation of the capability of various sounding-based parameters to discriminate among various classes of severe local storm environments; and, to explore the use of logistic regression as an alternative to a threshold-based analysis in assessing skill in inferring of tornado occurrence. Results suggest that certain sounding-based parameters show significant promise in their capability to discriminate among various classes of associated severe local storms. While this conclusion is similar to that found in previous investigations, it is clear from the relatively larger sample evaluated here that many parameter distributions show a station dependency across the study domain. In fact, both seasonal and spatial dependency may need to be addressed in some manner in order to avoid an analysis whose skill also various from station to station.

It should be noted that a formal regionalized analysis of sounding-based frequency distributions has not been conducted nor has a sensitivity analysis been carried out assessing the impact that the choice of base period has on the selection of predictors and on model skill. In addition, an evaluation of any variability in bias should be done to ensure that model skill does not vary greatly from year to year. Consequently, the results of the regionalization and logistic regression analysis presented in this report should be viewed as preliminary. Since logistic regression can also be used to model multiple possible outcomes, the results of such a “polychotomous” logistic analysis that models the probability of occurrence for each of the storm classes will be presented in a forthcoming report. Estimates of tornado days for other regions also will be reported.

6. REFERENCES

Brooks, H.E., C.A. Doswell III, and J. Cooper, 1994a: On the environments of tornadic and nontornadic mesocyclones. *Wea. Forecasting*, **9**, 606-618.

Craven, J.P., 2001: A baseline climatology of sounding derived parameters associated with deep, moist convection. Non-thesis research paper, School of Meteorology, University of Oklahoma, Norman, OK, 56 pp.

Darkow, J.M., 1969: An analysis of over sixty proximity soundings. *Preprints, 6th Conference on Severe Local Storms*, Chicago, IL, Amer. Meteor. Soc., 218-221.

Doswell, C.A. III, and D.W. Burgess, 1988: On some issues of United States tornado climatology. *Mon. Wea. Rev.*, **116**, 495-501.

Hosking, J.R.M. Hosking and J.R. Wallis, 1997: *Regional Frequency Analysis, An Approach Based on L-Moments*, Cambridge University Press, Cambridge, UK, 224 pp.

Hosmer, D.W., and S. Lemeshow, 2000: *Applied Logistic Regression, second edition*. John Wiley & Sons, Inc, New York, 375 pp.

Mead, C.M., 1997: The discrimination between tornadic and non-tornadic supercell environments: A forecasting challenge in the southern United States. *Wea. Forecasting*, **12**, 379-387.

Miller, R.C., 1972: Notes on analysis and severe-storm forecasting procedures of the air force global weather central, Technical Report 200 (Rev), Air Weather Service (MAC).

Rasmussen, E.N., and D.O. Blanchard, 1998: A baseline climatology of sounding-derived supercell and tornado forecast parameters. *Weather and Forecasting*, **13**, 1148-1164.

Wilks, D.S., 1995: *Statistic Methods in the Atmospheric Sciences*, Academic Press, 467 pp.